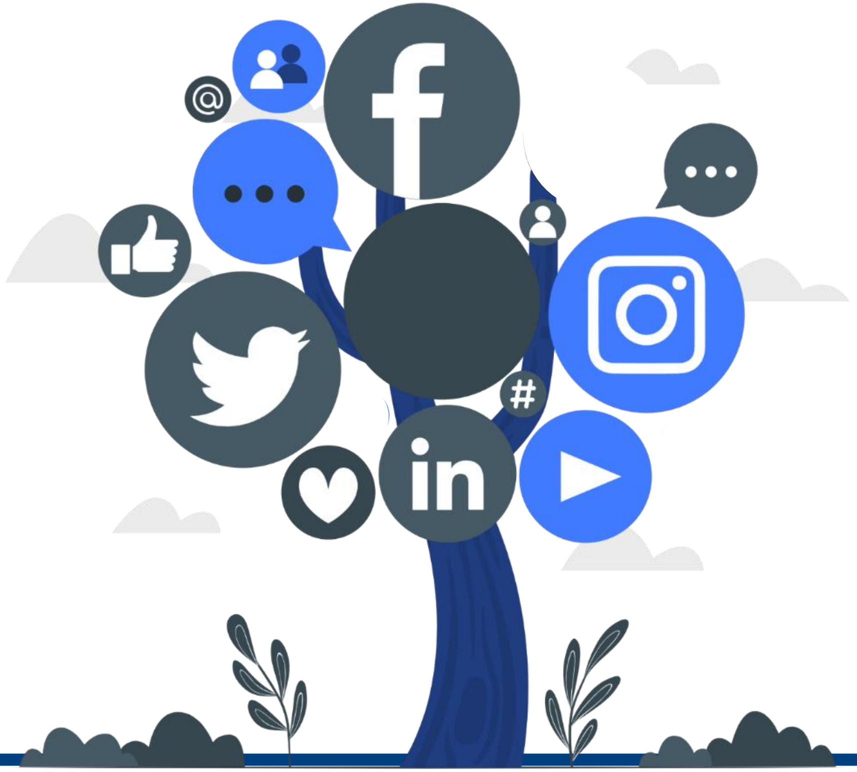
A comprehensive data-driven approach to optimizing Stanbic IBTC social media engagements across different platforms for enhanced Digital marketing



BY: ONABAJO MONSURAT

dicey tech

Stanbic Bank

INTRODUCTION

This project is the report of the dicey “Hack the feeds: Insights from social media data” data analysis hackathon project which aims to get actionable insights that could redefine the future of digital marketing

PROBLEM STATEMENT

Stanbic IBTC, despite having a vast presence on platforms like Facebook, Instagram, Twitter, and LinkedIn, struggles with inconsistent user engagement on their posts. While some content gets high interactions, others are largely overlooked, making it challenging to devise a consistent digital marketing strategy.

This inconsistency implies missed opportunities for meaningful customer interactions and potential inefficiency in resource allocation. To address this, a data-driven approach is essential to understand and enhance engagement factors, optimizing the bank's social media strategy for better marketing results.

OBJECTIVES

1. Delve deep into historical post metrics, identifying patterns and trends that underscore successful engagements.

2. Determine key distinguishing factors that set apart highly engaging posts from their less captivating counterparts.

3. Provide actionable recommendations to enhance StanbicIBTC's digital content strategy, with a focus on achieving consistent and amplified audience engagement.

4. Develop a machine learning model capable of predicting post engagement rates based on selected features, offering a predictive tool for future content strategies.

METHODOLOGY

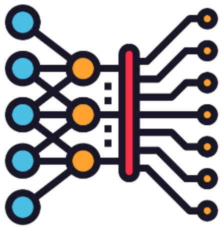
Data Gathering

Data Assessment

Data Cleaning

Feature Engineering

Data Visualization



Machine Learning

DATA GATHERING

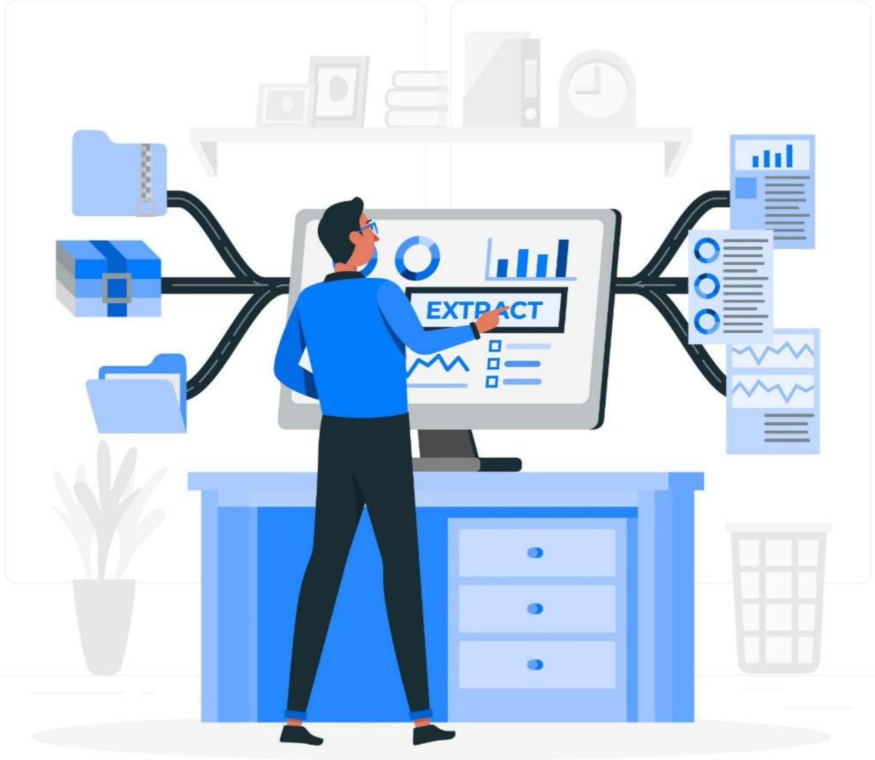
Source: The data used for this analysis was provided by Dicey Tech, encompassing social media metrics from Stanbic IBTC bank's engagements across four platforms: LinkedIn, Facebook, Instagram, and Twitter.

Format and Acquisition: The datasets, available in both CSV and XLS formats, were manually downloaded, ensuring the comprehensive capture of relevant social media metrics.

Data Assessment

Upon loading the datasets into a Jupyter Notebook environment, a preliminary examination revealed a consistent structure across the files with regard to columns. However, the row count varied, reflecting the different volumes of records for each platform.

There were columns with missing data and inconsistent datatypes. There is no duplicated rows in the data even though there are columns with duplicated values. This made me wonder why this was so. On further assessment I gained additional insights



DATA CLEANING

1. **Duplicates**: There were no duplicate records in all datasets.

2. Handling Missing Data: Missing data across all datasets were handled the similarly

a. A closer look revealed there are several columns with 100% missing data. While the proportion of these columns varied by dataset, they were uniformly removed for the sake of a more robust analysis. resulting in the following:

i. **Facebook:** 101 columns retained
ii. **Instagram:** 23 columns retained
iii. **Twitter:** 37 columns retained
iv. **LinkedIn:** 25 columns retained

b. Object-type columns with missing entries were filled with "Data Not Available" for clarity.

c. Rows containing NaN (Not a Number) and null values (Missing values) were dropped to ensure the integrity and accuracy of the analysis.

DATA CLEANING CONTINUES

3. Data Type Corrections: The columns were analyzed visually and programmatically. columns with misrepresented datatypes were cleaned the following ways.

i. Delimiters, such as commas and percentage signs, found in numeric columns, were removed for accurate computation.

ii. All columns were converted to their correct datatype. I ended up having datetime, integer, float, object and categorical columns

4. Time Frame Selection: To maintain consistency in temporal analysis, the data was filtered to include records from the years 2018 through 2023.

5. Outliers: Even though there were outliers in the data, they were not dropped as they might represent genuine data points.

6. Others: Columns containing uniform values were dropped as they would not add any insights to the analysis.

7. Empty rows in 'sent_by' columns was filled with 'Anonymous'

8. All rows with negative values was dropped

FEATURE ENGINEERING

Derived Columns and Their Rationale:

Columns Derived from date: Date and time when the post was published.

1. **Year:** Represents the year of the social media activity.
   - Rationale: Derived from the Date column to analyze year-over-year trends and assess the performance of posts across different years.

2. **Month:** Specifies the month of the social media activity.
   - Rationale: Useful for observing monthly patterns or seasonality in the data, aiding in understanding which months have higher or lower engagements.

3. **part_of_day_posted:** Categorizes the post by the time of day (morning, afternoon, evening).
   - Rationale: Posting time can influence visibility and engagement. This categorization aids in determining the best times to post for maximum impact.

4. **day_of_week**: Denotes the day of the week of the post.
   - Rationale: Certain days might see higher engagements due to user activity patterns. This column aids in pinpointing optimal days for posting.

5. **is_weekend:** Indicates if the post was made on a weekend.
   - Rationale: Weekends might have different user activity patterns. This binary indicator helps in contrasting weekday vs. weekend performance.

FEATURE ENGINEERING

Columns Derived from Tags

1. **Tags available?:** Indicates if the post has tags.
   - Rationale: Tags can significantly influence discoverability and engagement. This column helps in evaluating the impact of tags on post performance.

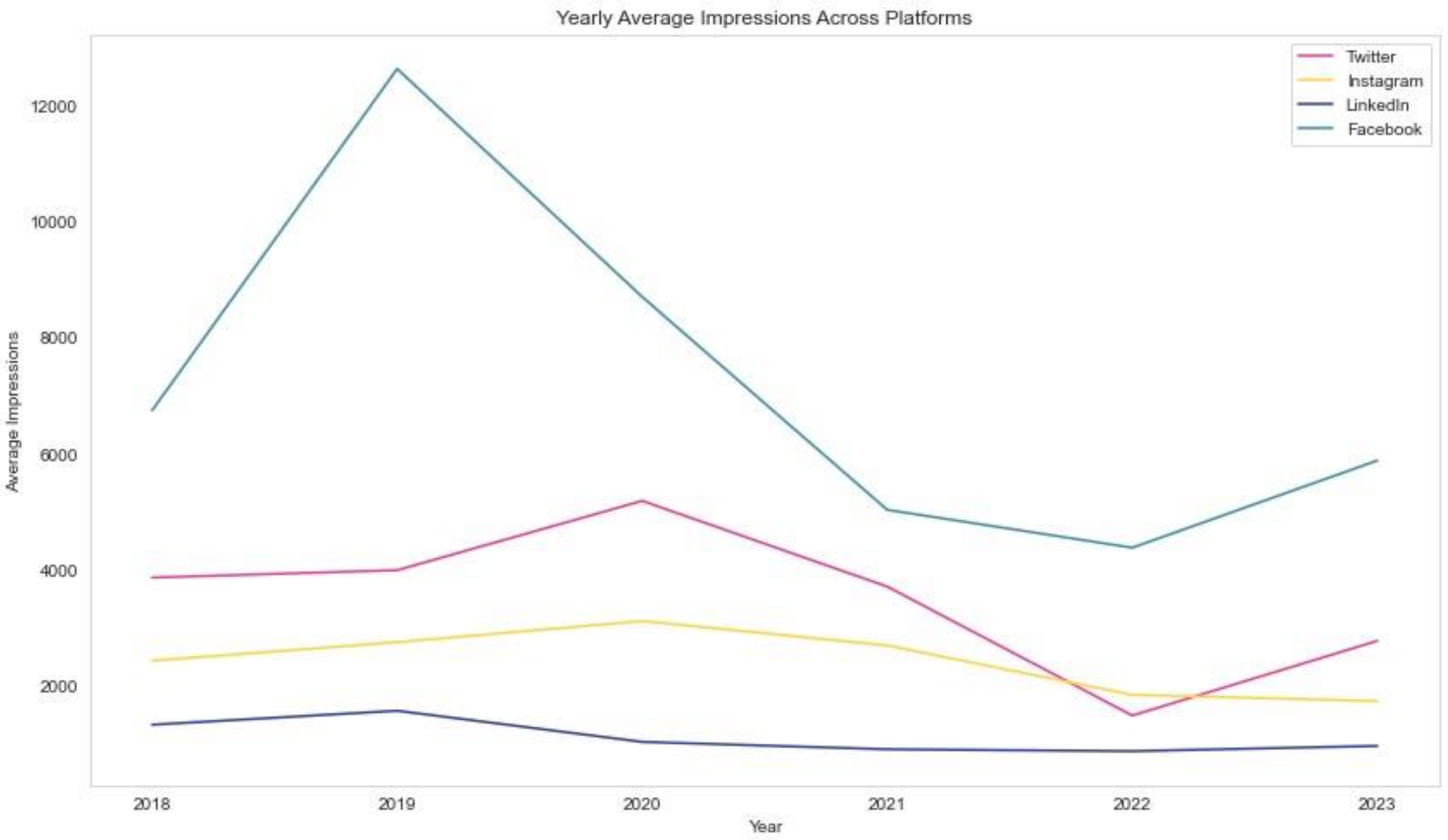
Columns Derived from Linked contents

1. **Linked Content(?):** Determines if the post contains linked content.
   - Rationale: Linked content can drive traffic to external sites, making it vital to assess its presence and potential correlation with engagement.

Columns Derived from Posts: Text content of the post.

1. **post_length:** Gives the length of the post content in characters.
   - Rationale: Post length can influence readability and engagement. This metric allows for analysis of optimal post lengths.
2. **contains_question**: Highlights if the post has a question.
   - Rationale: Questions can boost engagement by prompting user responses. This column assesses the impact of questions on post interactions.
3. **num_hashtags:** Counts the hashtags in the post.
   - Rationale: Hashtags can improve discoverability and categorization. Analyzing the number of hashtags helps in understanding their effect on post reach and engagement.

THE OBSERVED ENGAGEMENT TRENDS ACROSS MAJOR SOCIAL MEDIA PLATFORMS OVER RECENT YEARS



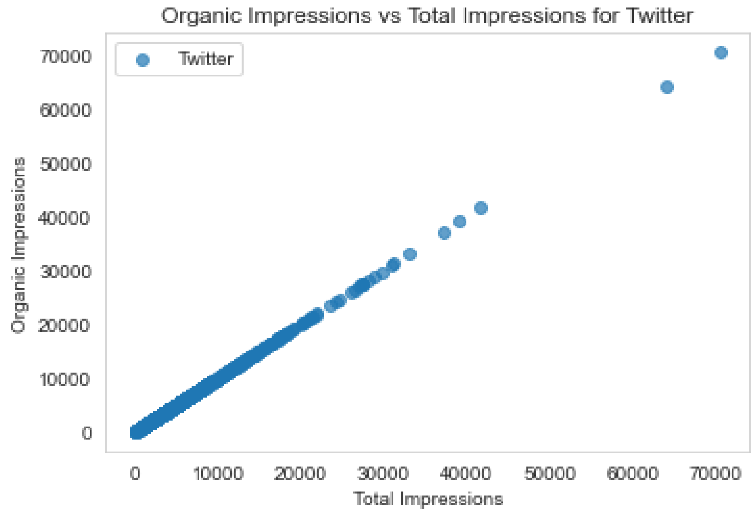
Observations

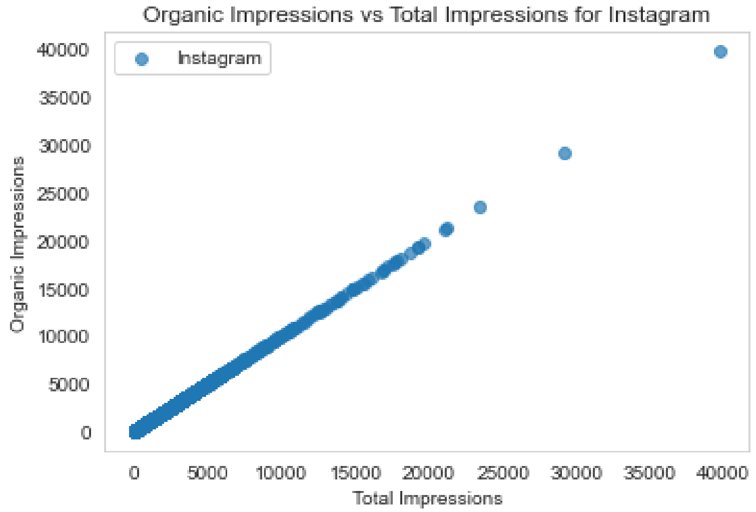
- Facebook emerges as the dominant social media platform, closely followed by Twitter which declined in 2022 but it is slowly gaining back its track.
- Impressions for all social media platforms consistently reach the thousands, indicating robust performance.

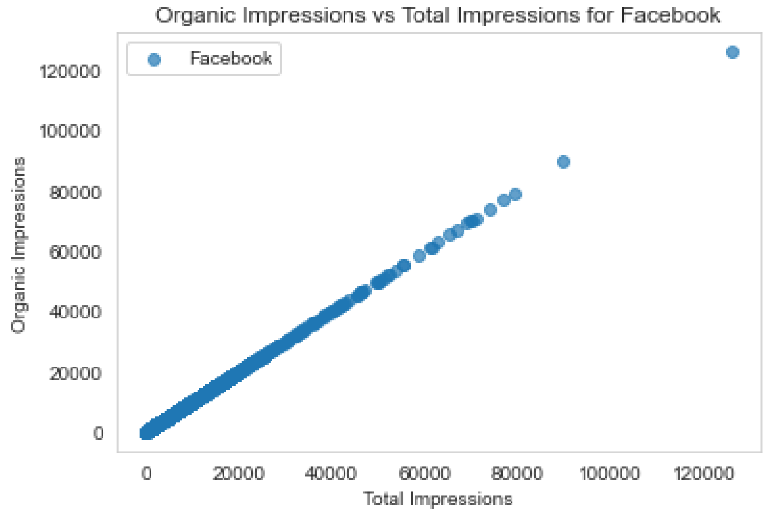
Key Insights:

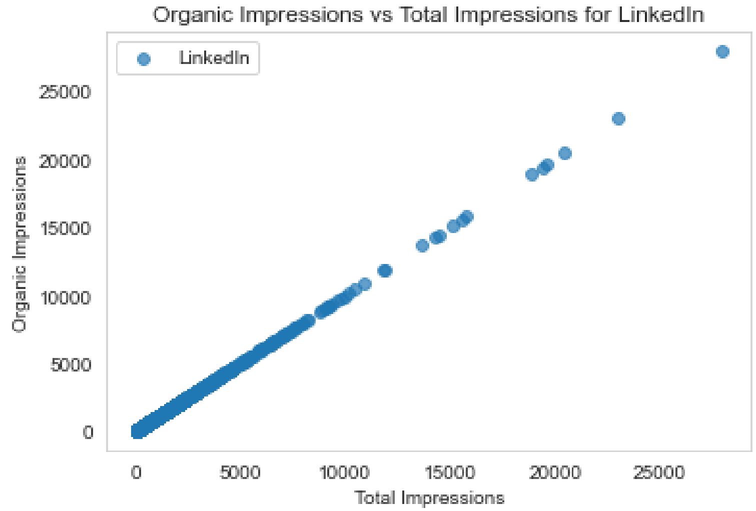
I noticed a peak in impression in different years across different datasets. This observation peaked my curiosity and lead me to further examine the correlation that existed between Organic Impressions and total impressions across all datasets. This allowed me ascertain if ever at some point it opted for the services of paid impressions vendors in any of its social media platform.

THE CORRELATION BETWEEN TOTAL IMPRESSIONS AND ORGANIC IMPRESSIONS ACROSS MAJOR SOCIAL MEDIA PLATFORMS FROM 2018 TO 2023.









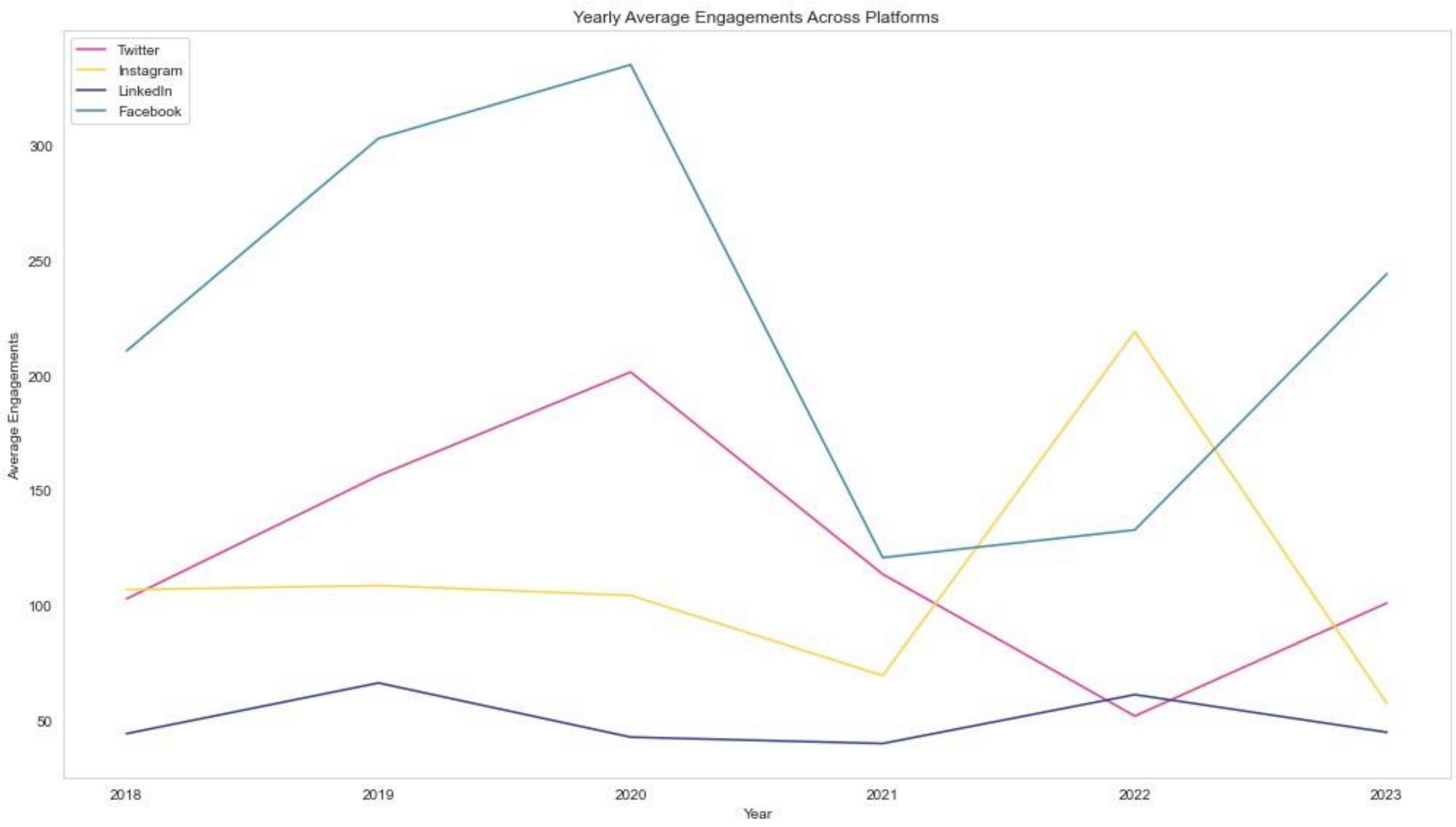
Observation:

- A pronounced linear correlation exists between Total Impressions and Organic Impressions. This suggests that from 2018 to 2023, the company relied solely on organic reach without investing in paid impressions on any of these platforms.

Key Insight:

- Now that I have ascertained that the company had not at any point opted for the services of paid impressions, I will be checking for yearly average engagements trends across all Social Media Platforms

THE OBSERVED ENGAGEMENT TRENDS ACROSS MAJOR SOCIAL MEDIA PLATFORMS OVER RECENT YEARS



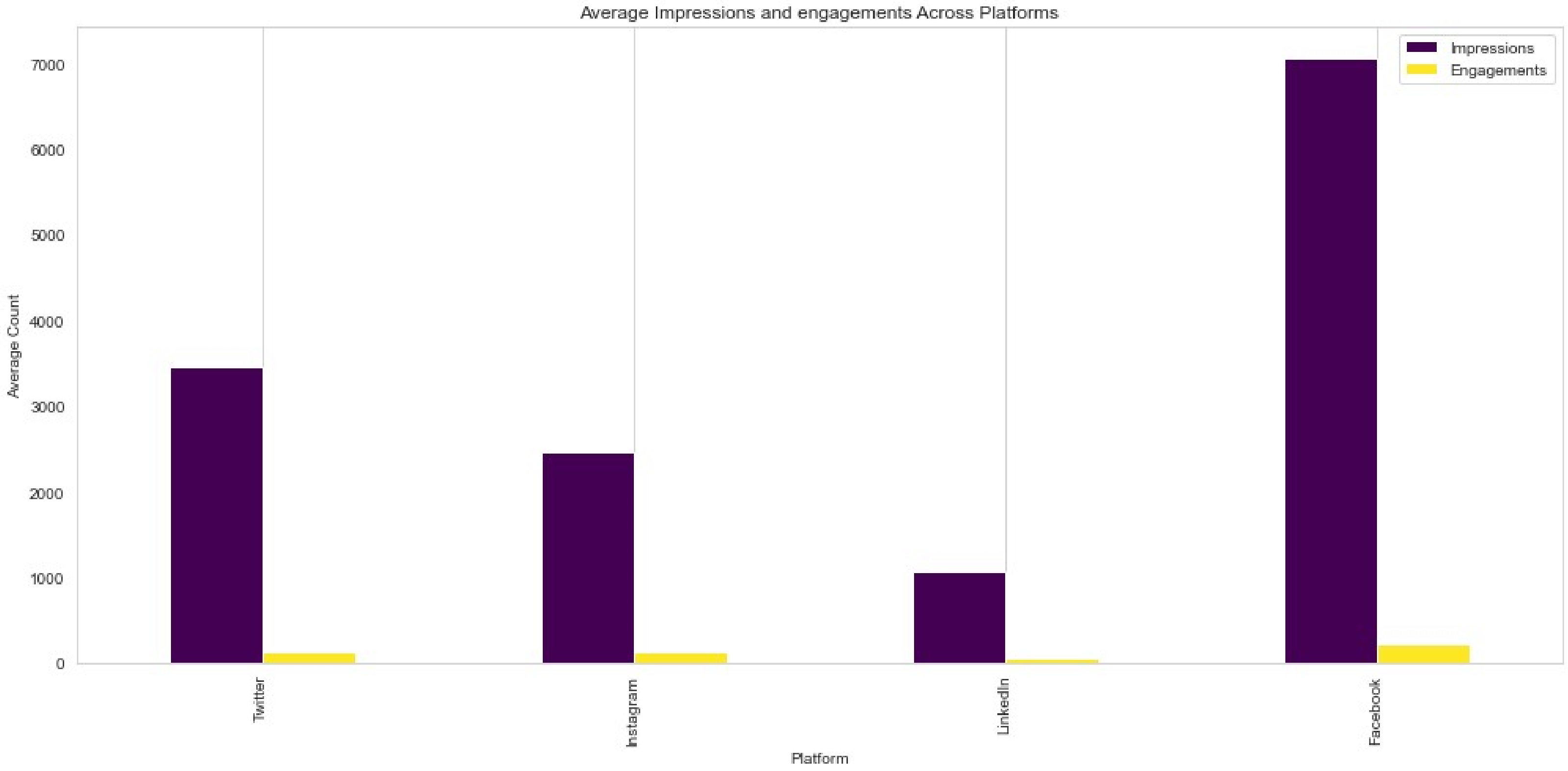
Observations:

- Up until 2021, Facebook led in engagements, followed by Twitter.
- However, post-2021, there was a notable decline in engagements across all platforms, potentially influenced by the workforce's return post-COVID-19 pandemic.
- Throughout the year, while impression rates remained high, actual engagements consistently lagged behind.

Next steps

Check for the relationship between Average Impressions and Average Engagements Across all social media platforms

ANALYZING THE DIFFERENCES BETWEEN IMPRESSION COUNTS AND ENGAGEMENT RATES ACROSS VARIOUS PLATFORMS.



Observations

Impressions: Facebook had the highest average impressions, followed by Twitter, Instagram, and Linkedin.

Engagements: Facebook also had the highest average engagements, with Instagram, Twitter, and Linkedin following.

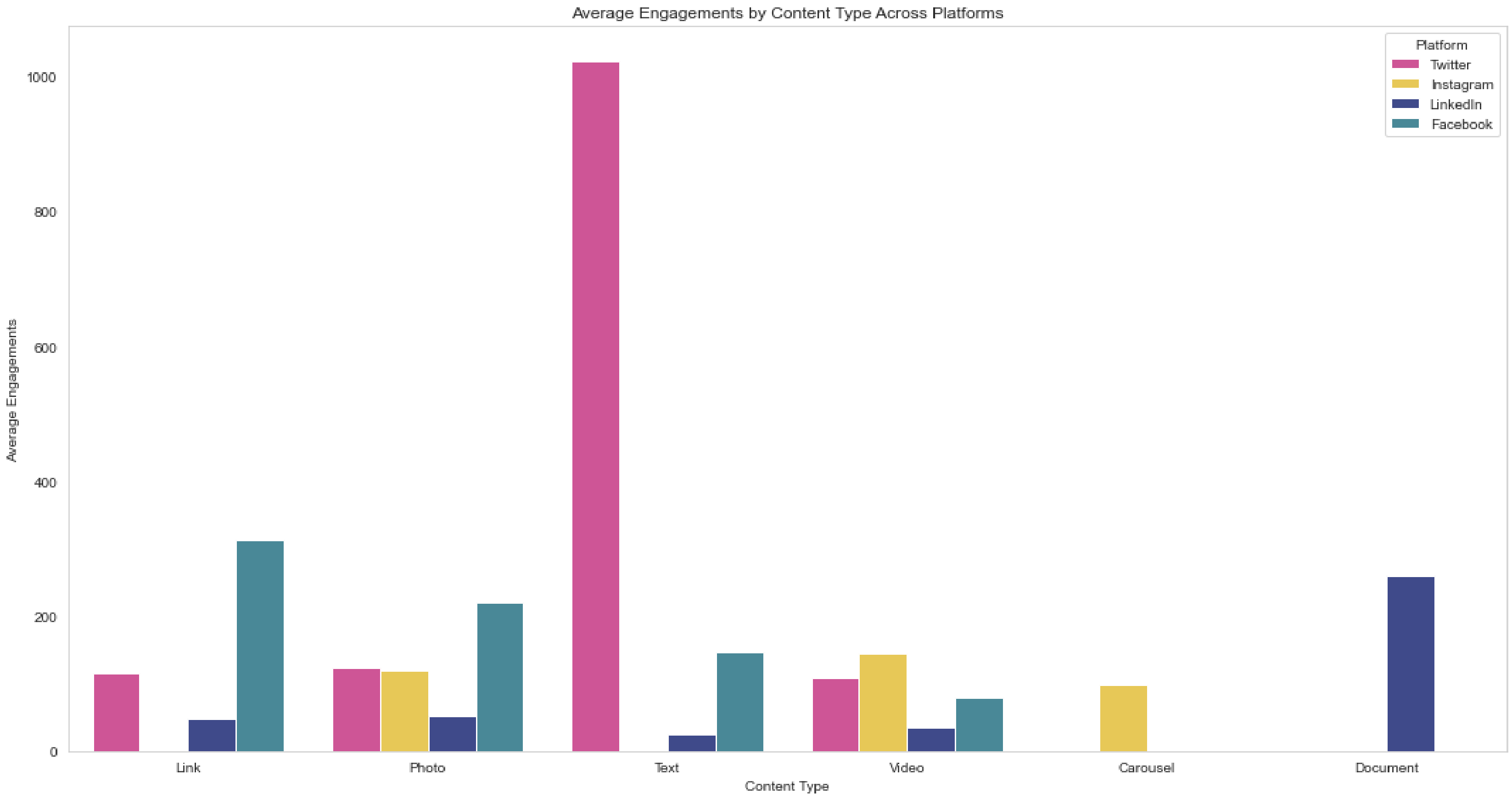
Next Steps

In order to Identify factors that influence high engagements with post, I will start by checking for the relationship between Content type and Average Engagements Across all social media platforms

Insights

Despite achieving high impression counts in the thousands, engagement rates fell short, not surpassing 10% of impressions. This gap between content visibility and user engagement suggests that the content, while viewed, might not be compelling to the audience. This trend could result in inefficient marketing investments and lost opportunities for genuine customer interactions.

THE RELATIONSHIP BETWEEN CONTENT TYPES AND ENGAGEMENTS ACROSS PLATFORMS



Observations

Twitter: Text received the highest average engagements, followed by photos. Links and text-only posts receive relatively lower engagements.

Instagram: Videos on Instagram receive higher average engagements than photos or Carousel.

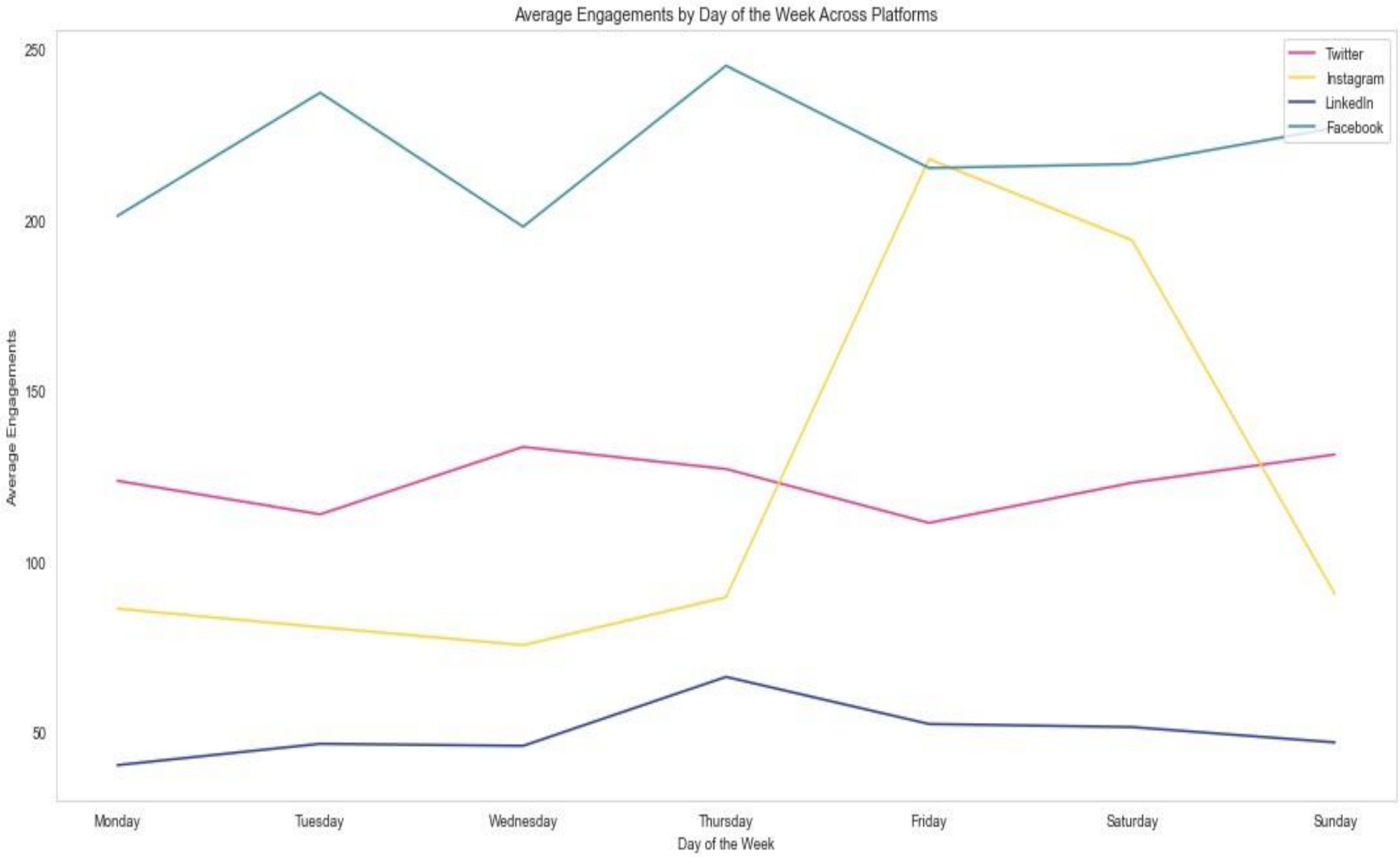
LinkedIn: Documents dominate in terms of average engagements on LinkedIn, followed by photos. Text-only posts and links receive comparatively lower engagements.

Facebook: Link lead in terms of average engagements on Facebook. Photos followed, while text and video received lower engagements.

Insights

Different content types (e.g., videos, images, texts) have varying engagement levels.
Videos garner higher engagement on platforms like Instagram, while articles might do better on LinkedIn because of the peculiarity of both platforms.
Aligning content type with platform preference can enhance engagement rates.

RELATIONSHIP BETWEEN AVERAGE ENGAGEMENTS BY DAYS OF THE WEEK ACROSS PLATFORMS



Observations

Twitter: Engagements on Twitter are fairly evenly distributed across the weekdays, with a slight peak on Wednesday.

Instagram: Instagram Engagements peak around mid-week, especially on Wednesdays- Fridays. The engagement frequency is lower during the weekends.

LinkedIn: LinkedIn sees a high engagement trend during the weekdays with a huge peak on Thursdays. Engagements decline significantly over the weekends, which is expected for a professional networking platform.

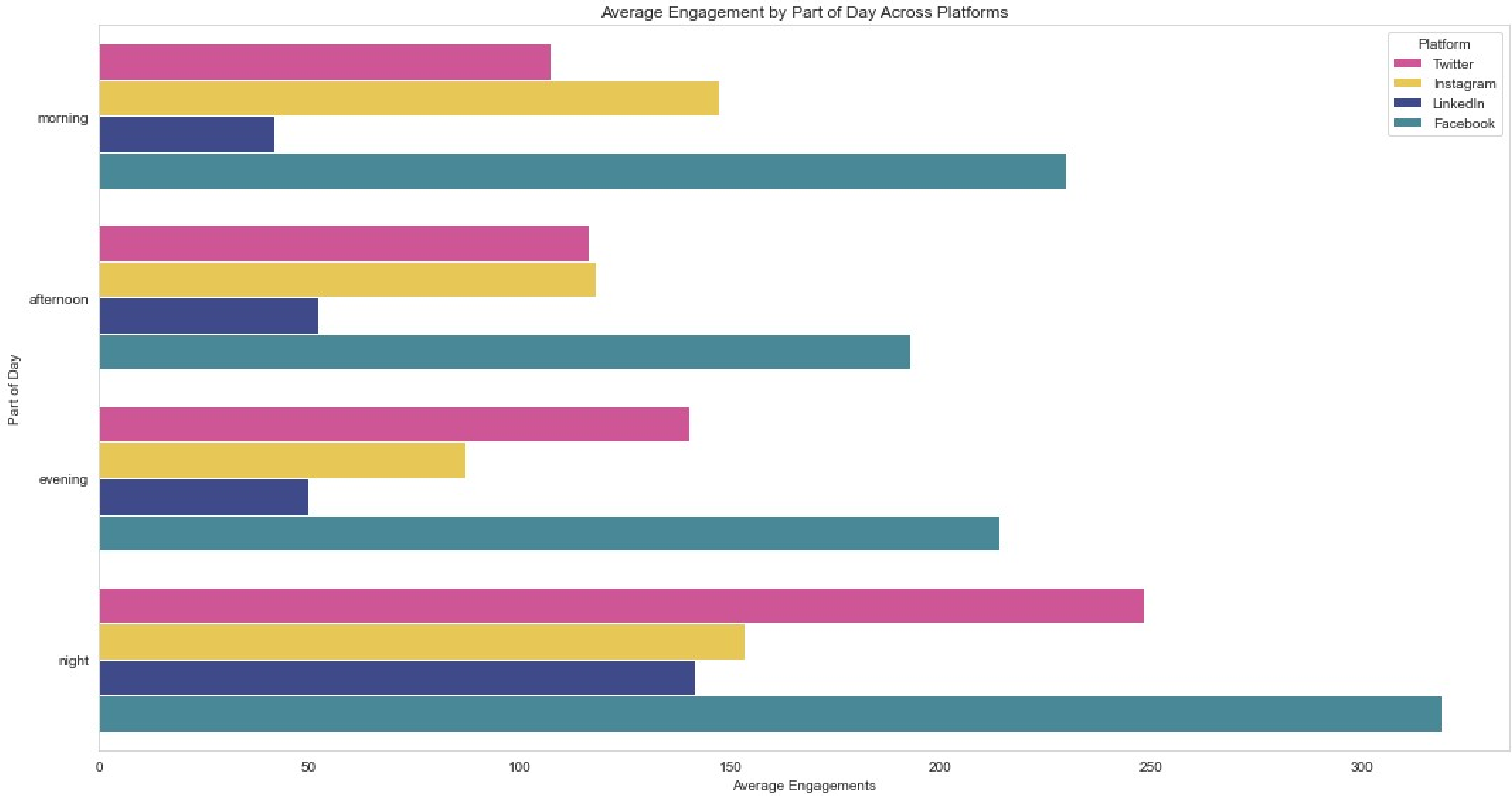
Facebook: Facebook has the highest engagements across platforms a relatively even engagements of posts across all days, with a slight dip on Wednesdays.

Insights

Engagement patterns vary by weekdays, with certain days showing higher interactions than others.

This trend can guide content scheduling to capitalize on days with naturally higher engagement tendencies.

REPORTS ON AVERAGE ENGAGEMENT BY PART OF DAY ACROSS PLATFORMS



Observations

Post made at night between 9pm-12am and in the morning between 5am-12pm usually tends to have higher engagements.

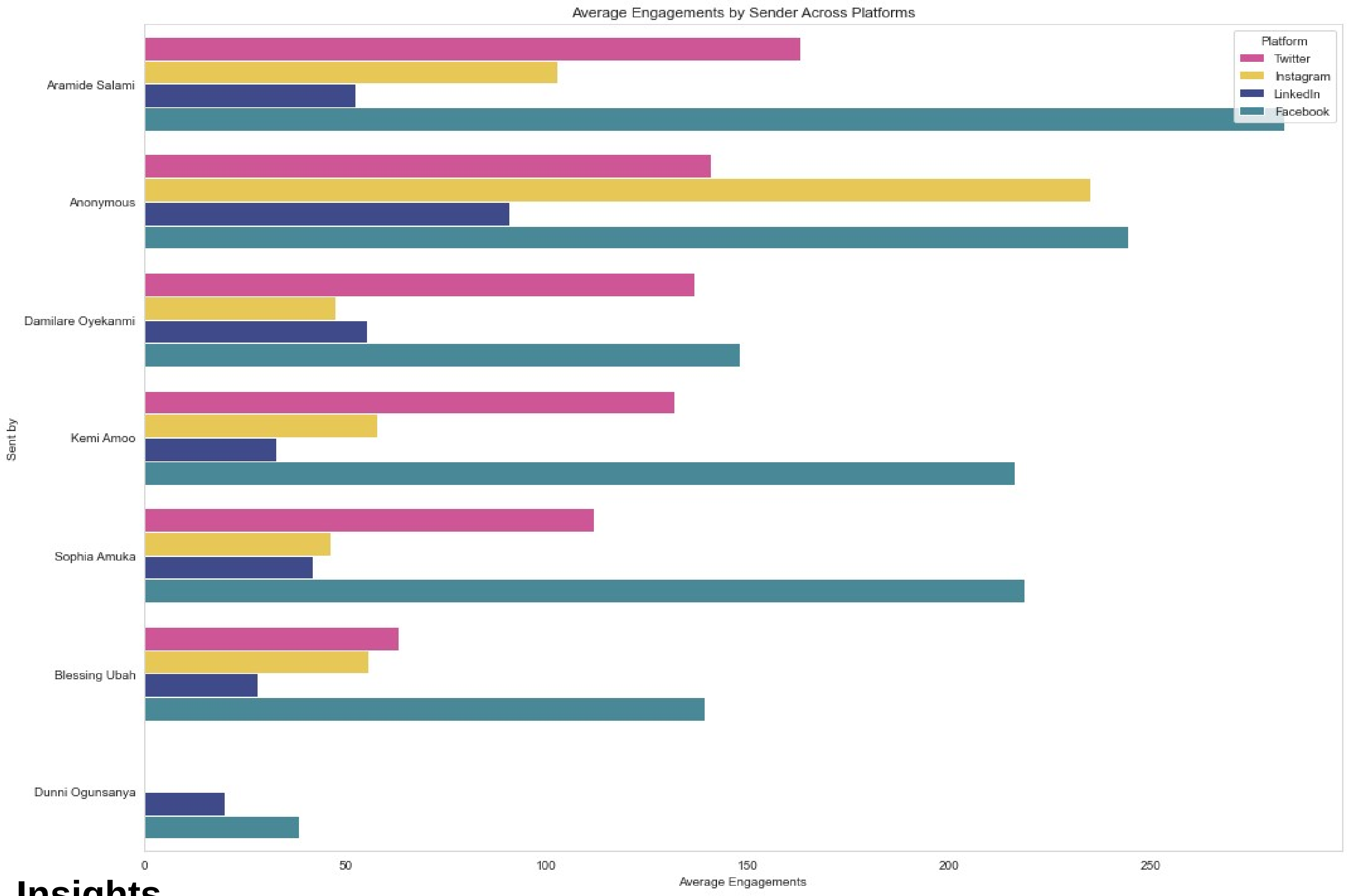
Insights

- Time of posting plays a crucial role in engagement.
- Certain times of the day, such as nights and morning drive more interactions on all platforms.
- Understanding these peak times can aid in scheduling content for maximum impact.

Next Steps

Check for the relationship between People who sent the post and The post average Engagements Across all social media platforms

REPORTS ON AVERAGE ENGAGEMENTS BY SENDER ACROSS PLATFORMS



Observations

Here are the average engagements for top 3 sender across the platforms:

Twitter:
- Aramide Salami
- Anonymous
- Damilare Oyekanmi

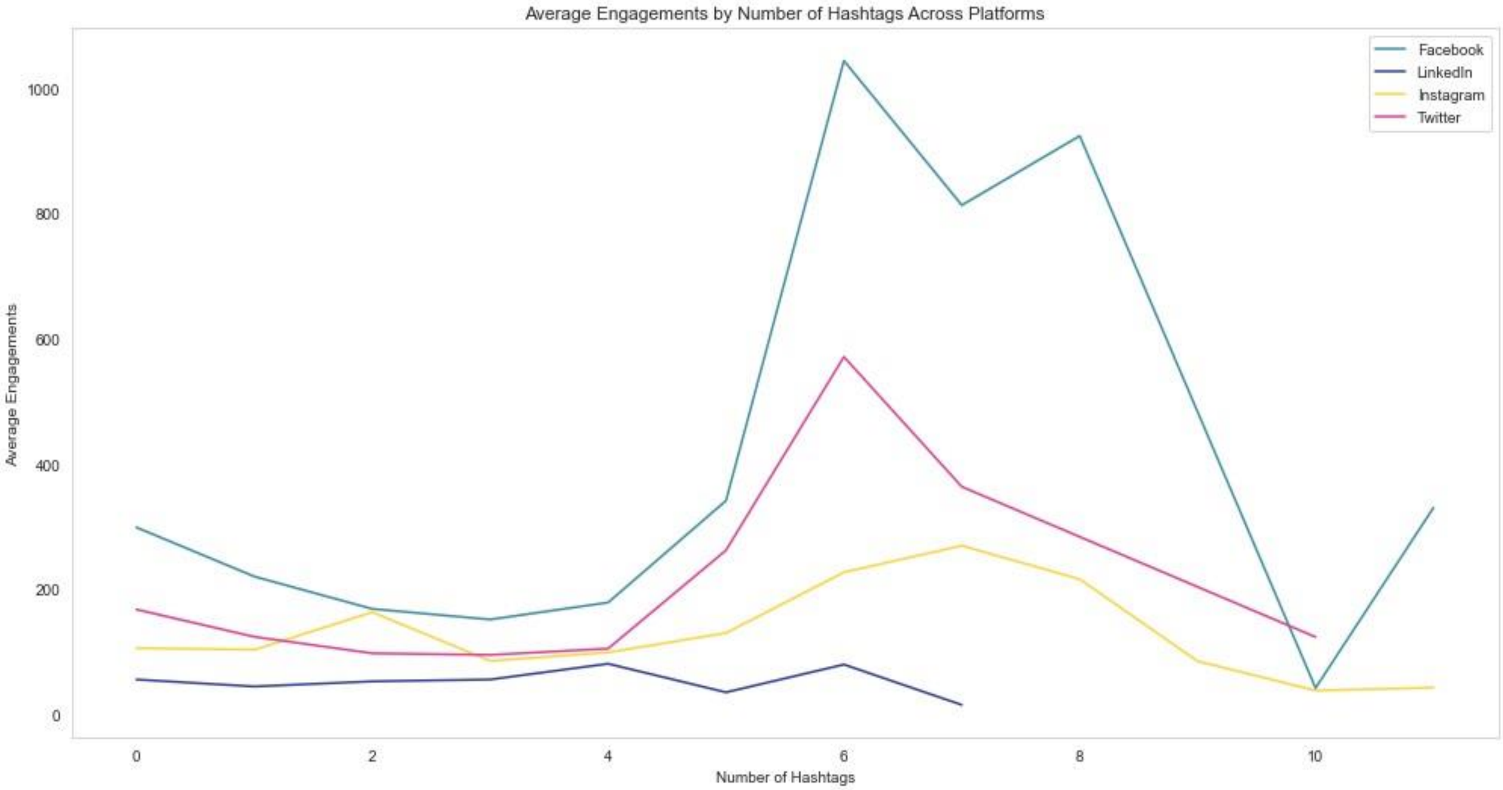
Instagram:
- Anonymous
- Aramide Salami
- Kemi Amoo

LinkedIn:
- Anonymous
- Damilare Oyekanmi
- Aramide Salami

Facebook:
- Aramide Salami
- Anonymous
- Sophia Amuka

Insights
- Engagement levels were influenced significantly by the sender or author of the content across all platforms.
- Some senders like Aramide Salami and Anonymous (The name was empty so it was replaced with 'Anonymous') consistently achieve higher engagements, indicating their content resonates more with audiences.
- Senders with lower engagement rates such as Dunni Odusanya might benefit from content strategy training.

AVERAGE ENGAGEMENTS BY NUMBER OF HASHTAGS ACROSS PLATFORMS ANALYSIS AND REPORT



Observations

- **Twitter**: Engagements seem to rise with an increase in the number of hashtags, peaking around 4-6 hashtags.
- **Linkedin**: There's a peak at around 3-4 hashtags, followed by a decline.
- **Facebook**: Average engagements seems to rise when hashtags are between the range of 4-6.
- **Instagram**: While not focusing on the outliers in the dataset, Average engagements seems to rise when hashtags are between the range of 5-8.

Insights

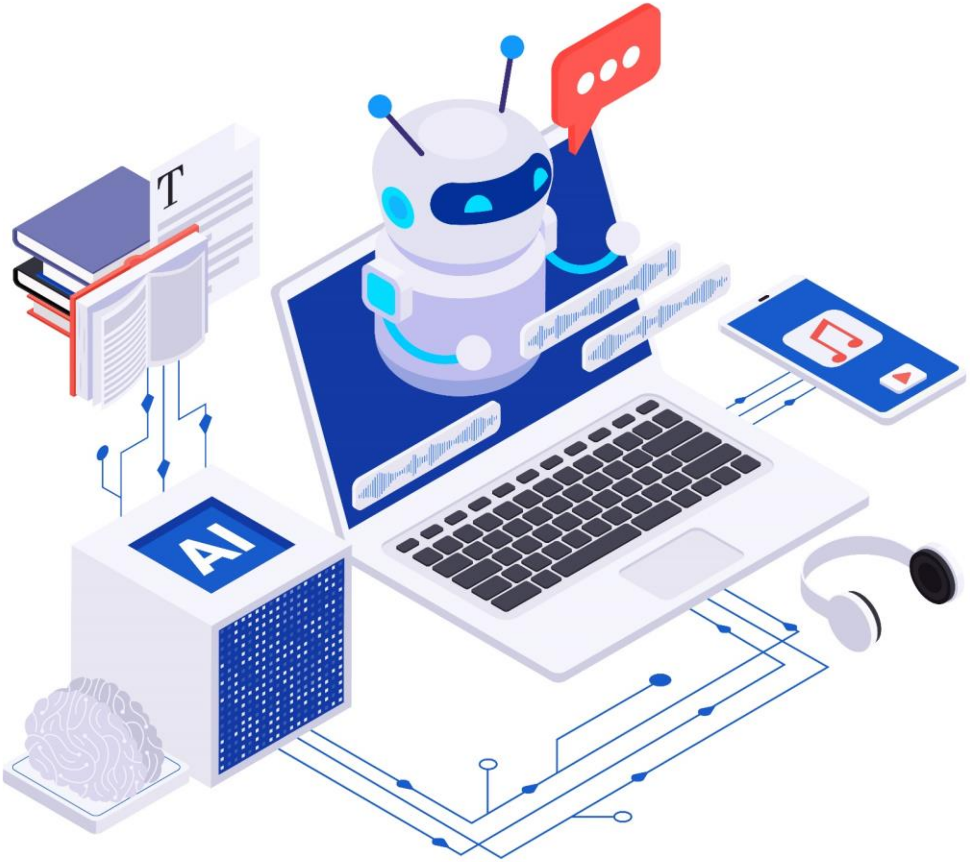
- There is a correlation between the number of hashtags used and engagement levels across platforms. Lower hashtags were observed to yield higher engagements.
- Using an excessive number of hashtags might reduce engagement rates across all platforms.
- Optimal hashtag usage varies by platform; for instance, Twitter showed peak engagement at a lower number of hashtags than Instagram.

MACHINE LEARNING TRAINING, TESTING AND EVALUATION METRICS

From four datasets, I identified 10 features consistently present post-2019 to build my machine learning model. These features include: Sent by, day_of_week, Month, is_weekend, contains_question, Content Type, part_of_day_posted, num_hashtags, Network, and Year, with 'Engagements' being my key predictive target.

These features were chosen due to their accessibility before any post is made. My main goal is to predict the engagement rate of content before it's published, providing insights into potential post performance.

The datasets were merged, preprocessed, and primed for machine learning. Object-type columns transitioned to numerical representations, and outliers were excluded. A baseline model was trained and it was scored using both Mean Absolute Error and R^2 Score metrics.



MACHINE LEARNING TRAINING, TESTING AND EVALUATION METRICS

Metrics Defined

- **Mean Absolute Error (MAE):** Represents the average of the absolute differences between predicted and actual values. A lower MAE suggests a better-fitting model.
- **R^2 Score (Coefficient of Determination):** Indicates the proportion of the variance in the dependent variable that is predictable from the independent variables. It ranges from 0 to 1, with 1 indicating perfect predictions and 0 meaning the model does no better than simply predicting the mean of the target variable.

Model Performance:

- **Baseline Model**:
  - MAE: 103
  - R^2 Score: 0.0

- **Random Forest Classifier:** The model exhibits strong performance on the training dataset:
  - MAE: 0.35
  - R^2 Score: 0.799

  - But its effectiveness diminishes on the test dataset
  - R^2 Score: 0.292

MACHINE LEARNING TRAINING, TESTING AND EVALUATION METRICS

- To address the model under fitting, I explored more complex and powerful models and got the following R^2 Score:
  - Linear Regression: 0.368
  - Ridge Regression: 0.368
  - Random Forest: 0.294
  - Gradient Boosting: 0.391
  - KNN: 0.301
  - Support Vector Regressor: 0.389

- I then tuned the highest performing model being the Gradient Boosting Regressor:
  - MAE: 0.65
  - R^2 Score: 0.408

Note

While the Gradient Boosting Regressor model showed remarkable improvement from the baseline model, its performance still fell short of expectations. Future steps will involve gathering more data to further refine and optimize the model.

ACTIONABLE RECOMMENDATIONS FOR ENHANCING ENGAGEMENT RATES OF STANBIC IBTC POSTS

Platform Strategy:

- Continue leveraging Facebook as the primary platform while recognizing the resurgence of Twitter post-2022.
- Allocate resources to platforms based on their performance and growth trends.

Optimizing Impressions & Engagements:

- Given the consistent high impression rates but lower engagement, company should focus on content that resonates and compels the audience to engage.
- They should consider experimenting with paid impressions to augment organic reach and measure its impact on engagements.

Content-Type Strategy:

- **Instagram**: They should prioritize creating video content.
- **LinkedIn:** They should prioritize producing and sharing more document-based content.
- **Facebook:** They should prioritize developing and sharing linked posts.
- **Twitter:** They should prioritize sharing text and image combinations texts.

ACTIONABLE RECOMMENDATIONS FOR ENHANCING ENGAGEMENT RATES OF STANBIC IBTC POSTS

Engagement Timing Strategy:

- **LinkedIn:** Schedule key posts for weekdays, especially Thursdays.
- **Instagram:** Focus on mid-week posts, from Wednesday to Friday.
- **All Platforms:** Optimize post timings for mornings (5 am-12 pm) and nights (9 pm-12 am) to capture maximum audience attention.

Content Creator Strategy:

- Encourage top-performing senders like Aramide Salami and 'Anonymous' to share insights and best practices with the team.
- Training and Development: Organize content creation workshops and strategy sessions for senders with lower engagement rates to improve their output.

Hashtag Strategy:

- **Optimal Hashtag Use:** Limit the number of hashtags based on platform insights:
    - Twitter: 4-6 hashtags.
    - LinkedIn: 3-4 hashtags.
    - Facebook: 4-6 hashtags.
    - Instagram: 5-8 hashtags.
- Conduct regular research to identify trending and relevant hashtags that resonate with the target audience.

FUTURE WORK

Data Collection:

- Historical Data: Retrieving older post data for all platforms will make us to understand long-term trends in all platforms.
- User Demographics: Understanding who engages with the content can offer deeper insights into tailoring content for specific audience segments.

Machine Learning Model:

- With an expanded dataset, a more refined machine learning model that more accurately predicts the engagement rate of posts can be developed. This predictive model would utilize features available prior to posting, providing insights into:
  - **Expected Engagement:** Before publishing a post, the team can have an idea of how it might perform, allowing for real-time adjustments.
  - **Feature Importance:** Understanding which features (like time of posting, hashtags, content type) most influence engagement can guide content strategy.
  - **Scenario Testing:** The model could be used to test different post scenarios and select the one predicted to have the highest engagement.

Integration of Predictive Models with Content Strategy:

- **Content Planning:** Utilize the predictive model in content planning sessions. By predicting engagement rates, the content team can prioritize certain types of posts or adjust strategies based on predicted outcomes.
- **A/B Testing:** Use the model to test different content versions and optimize based on predicted engagement.

IMPLEMENTATION PATH

1. Content Audit: Review existing content to identify what's working and areas of improvement.
2. Training Workshops: Organize monthly training for content creators, focusing on platform-specific best practices as recommended in the actionable insights slide for creating engaging contents for audience.
3. Analytics and Feedback Loop: Set up regular review cycles (e.g., monthly) to assess the impact of implemented strategies and iterate based on user engagement with content.
4. Engage the Audience: Consider surveys or polls to gather direct feedback from the audience regarding content preferences and areas of improvement.

CONCLUSION

By focusing on these future directions, Stanbic IBTC can harness the power of data and machine learning to optimize its social media strategy, ensuring content not only reaches its audience but also resonates and engages them. Predicting engagement rates in advance will allow for a more proactive approach to content creation and scheduling, ultimately leading to improved digital marketing outcomes.